

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-318905

(43)Date of publication of application : 16.11.2001

(51)Int.Cl.

G06F 15/177

G06F 3/06

G06F 12/00

G06F 15/00

G06F 15/167

(21)Application number : 2000-133585

(71)Applicant : MATSUSHITA ELECTRIC IND CO LTD

(22)Date of filing : 02.05.2000

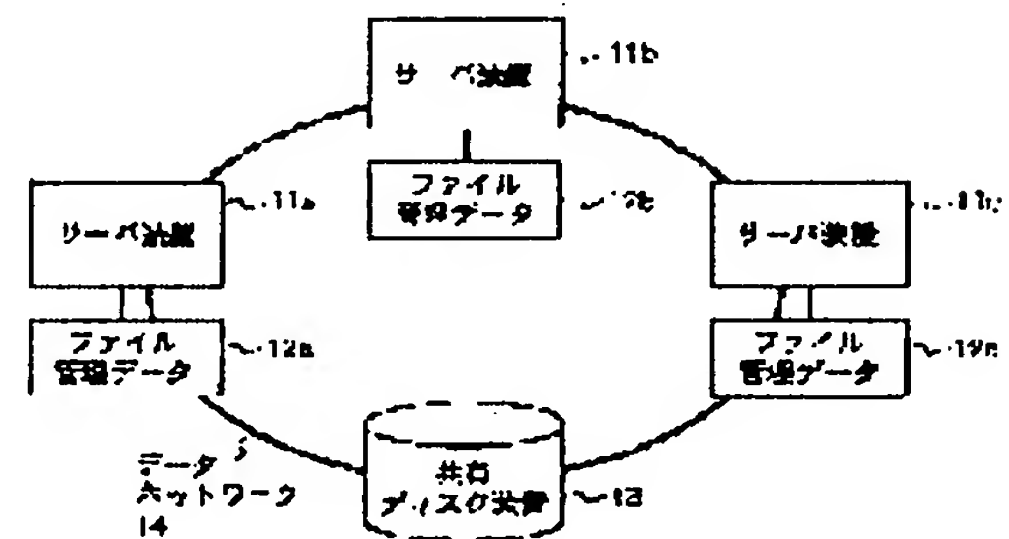
(72)Inventor : NAKATSUKA MONTA

(54) DISK SHARED TYPE DISTRIBUTED SERVER SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To solve problems that the installation cost of another mechanism for executing exclusive control is required though each server device executes the exclusive control by the mechanism existing in a place (e.g. a device annexed to a shared disk or a completely different device) different from the server device concerned in a distributed file system, and when a fault is generated in the other mechanism, the whole system is stopped.

SOLUTION: The system for sharing a disk by plural server devices divides information written in a file management data and each server device independently manages the information. Only when a request is not satisfied only by the file management data stored in the server device itself, the server device inquires the other server device whether the device has the information of the file management data or not.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号
特開2001-318905
(P2001-318905A)

(43)公開日 平成13年11月16日(2001. 11. 16)

(51)Int.Cl. ⁷	識別記号	F I	テ-マ-ト*(参考)
G 0 6 F 15/177	6 8 2	G 0 6 F 15/177	6 8 2 B 5 B 0 4 5
3/06	3 0 1	3/06	3 0 1 J 5 B 0 6 5
12/00	5 2 0	12/00	5 2 0 J 5 B 0 8 2
	5 4 5		5 4 5 A 5 B 0 8 5
15/00	3 1 0	15/00	3 1 0 A

審査請求 未請求 請求項の数 9 O L (全 9 頁) 最終頁に続く

(21)出願番号 特願2000-133585(P2000-133585)

(22)出願日 平成12年 5 月 2 日(2000. 5. 2)

(71)出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72)発明者 中塚 紋太

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(74)代理人 100081813

弁理士 早瀬 憲一

Fターム(参考) 5B045 BB12 BB28 BB47 DD03 DD16

EE03 EE25 GG01

5B065 BA01 CA02 CC03 CC08 CE26

ZA01 ZA08 ZA15

5B082 CA08 HA08 HA09

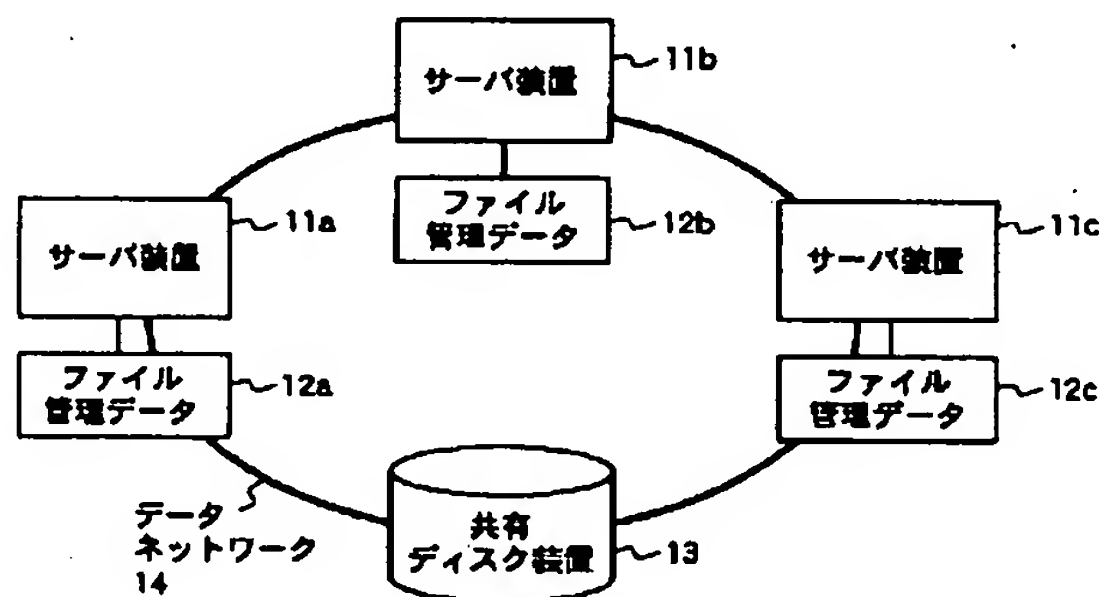
5B085 BA06 BE07 BG04 BG07

(54)【発明の名称】 ディスク共有型分散サーバシステム

(57)【要約】

【課題】 分散型ファイルシステムでは、サーバ装置とは別の場所（例えば、共有ディスクに付属した装置、または全く別の装置）において、排他制御を司る機構が存在し、各々のサーバ装置はその機構に頼って排他制御を行う。これによりファイルアクセスの負荷は分散されるが、排他制御を司る別機構を設けるコストがかかる。また上記別機構に障害が発生した場合、システム全体が停止してしまうなどの問題があった。

【解決手段】 複数のサーバ装置でディスクを共有するようなシステムにおいて、ファイル管理データに書かれた情報を分割し、サーバ装置はその情報を独立して管理する。サーバ装置は、クライアントからの要求に従い、自らが持つファイル管理データだけでは要求が満たせない場合のみ、他のサーバ装置にファイル管理データの情報を問い合わせる。



【特許請求の範囲】

【請求項1】 データの入出力を目的とした複数サーバ装置が、データネットワークを介して、データを格納するためのディスク装置を共有する構成をとるディスク共有型分散サーバシステムにおいて、

前記ディスク装置におけるデータを論理的なファイルとして扱うためのファイル管理データを、予め2つ以上の情報に分割しておき、前記分割された情報を、各サーバ装置内部において管理するファイルシステムを備え、
前記複数サーバ装置のうちの1つがこれに接続されるクライアントからのデータ配信要求を受けると、該サーバ装置のファイルシステムは、管理するファイル管理データの情報の中から該当するコンテンツファイルを検索し、該当コンテンツファイルが存在した場合、前記ディスク装置に対してアクセスを行ない、読み出しを行う、ことを特徴とするディスク共有型分散サーバシステム。

【請求項2】 請求項1記載のディスク共有型分散サーバシステムにおいて、

前記該当コンテンツファイルが存在しなかった場合、前記ファイルシステムは、前記データネットワークを介して接続されている他のサーバ装置に対して、前記該当コンテンツファイルが存在するかどうかを確認し、前記該当コンテンツファイルの情報を持つ前記他のサーバ装置よりデータの読み出しに必要な情報を受け取る手段を備え、

前記ディスク装置に対してアクセスを行い、読み出しを行う、

ことを特徴とするディスク共有型分散サーバシステム。

【請求項3】 請求項1または請求項2記載のディスク共有型分散サーバシステムにおいて、

前記複数サーバ装置のうちの1つがこれに接続されるクライアントからデータ記録要求を受けると、ファイルシステムは前記データネットワークを介して接続されている、他のサーバ装置が管理するファイル管理データの情報の中に、記録要求されているコンテンツファイルと同一名のものがない事を確認する手段と、

前記ファイルシステムが管理するファイル管理データの情報の中から、記録するのに必要なディスク容量を確保する手段を備え、

前記ディスク装置に対して書き込みを行う、

ことを特徴とするディスク共有型分散サーバシステム。

【請求項4】 請求項3記載のディスク共有型分散サーバシステムにおいて、

前記ファイルシステムがファイル管理データを参照して記録するのに必要なディスク容量が確保できなかった場合、

前記ファイルシステムは、前記データネットワークを介して接続されている、他のサーバ装置内部にあるファイルシステムに対してコンテンツファイルの情報の管理を要求する手段と、

前記ディスク装置に対して書き込む際に、前記コンテンツファイルの情報の管理を要求した他のサーバ装置よりファイルアクセス（書き込み）に必要な情報を受け取る手段を備え、

前記ディスク装置に対して書き込みを行う、

ことを特徴とするディスク共有型分散サーバシステム。

【請求項5】 請求項3または請求項4記載のディスク共有型分散サーバシステムにおいて、

各サーバ装置内部のファイルシステムは、共有している前記ディスク装置に記録されている全コンテンツファイル名および前記コンテンツファイルの情報を管理しているサーバ装置を対応付けさせた管理テーブルを持つ、

ことを特徴とするディスク共有型分散サーバシステム。

【請求項6】 請求項5記載のディスク共有型分散サーバシステムにおいて、

コンテンツファイルの追加または削除を行ったサーバ装置は、前記データネットワークを介して接続されている他のサーバ装置に対して、前記管理テーブルの更新を指示する手段を備えた、

ことを特徴とするディスク共有型分散サーバシステム。

【請求項7】 請求項5または請求項6記載のディスク共有型分散サーバシステムにおいて、

各サーバ装置内部のファイルシステムは、前記分割した記録領域における空き情報を空き領域テーブルとして持つ、

ことを特徴とするディスク共有型分散サーバシステム。

【請求項8】 請求項7記載のディスク共有型分散サーバシステムにおいて、

前記空き領域情報の更新を行ったサーバ装置は、前記データネットワークを介して接続されている他のサーバ装置に対して、前記空き領域テーブルの更新を指示する手段を備えた、

ことを特徴とするディスク共有型分散サーバシステム。

【請求項9】 請求項3または請求項4記載のディスク共有型分散サーバシステムにおいて、

システム初期化時に、前記ディスク装置における記録領域を論理的に分割する手段と、

前記分割した記録領域に内在するコンテンツファイルの情報を取得する手段とを備えた、

ことを特徴とするディスク共有型分散サーバシステム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、データの入出力を目的とした複数サーバ装置が、データネットワークを介して、データを格納するためのディスク装置を共有する構成をとるディスク共有型分散サーバシステムに関するものである。

【0002】

【従来の技術】 サーバ装置とは、例えばクライアントからのデータ配信要求に応じて、ディスク装置からデータ

を読み出しクライアントに送信するというものである。このサーバ装置内部において、ディスク上に格納したデータの集合を論理的なファイルとして扱うため、独自のファイル管理データによりファイルを管理するのが、ファイルシステムと呼ばれるものである。

【0003】このようなサーバ装置を分散型モデルによって実現する上では、複数のサーバ装置間で、このファイル管理データの排他制御を如何にして行うかが課題であった。

【0004】従来では、あるサーバ装置内部に排他制御を司る機構がファイル管理データを一括して管理することで、システム全体のファイルオペレーションの承認を1つのサーバ装置が行っていた。この方式では実装やデータの一貫性という面に優れている反面、サーバ装置の台数が増える事により排他制御機構によるボトルネックが発生してしまい、また排他制御を行うサーバ装置に故障が発生するとシステム全体に影響を与えてしまう。

【0005】例えば特開平11-146343号公報に開示されたディスク共有システムに記載されているように、排他制御を司る機構をシステムの一部に設け、ファイルオペレーションに関する承認を一手に引き受ける事で、ファイル管理データの一貫性を保つというものである。

【0006】

【発明が解決しようとする課題】しかしながらこのような方式では、排他制御をサーバ装置とは別機構にすることでコスト面の負荷があり、また排他制御機構がシステムの弱点となり、そこで故障が発生するとシステム全体に影響を与えることとなる。このため、排他制御機構部には、信頼性向上のための機構が別途必要となり、さらにコストの負荷が増える。

【0007】本発明は、ディスク装置を共有する構成をとるディスク共有型分散サーバシステムにおいて、特定のサーバ装置に対して負荷を集中させることなく、またコストを低く抑える事ができ、かつ障害発生時にシステム全体に致命的な影響を与えることのないファイルシステムである、ディスク共有型分散サーバシステムの提供を目的としている。

【0008】

【課題を解決するための手段】この課題を解決するために、請求項1に記載のディスク共有型分散サーバシステムは、データの入出力を目的とした複数サーバ装置が、データネットワークを介して、データを格納するためのディスク装置を共有する構成をとるディスク共有型分散サーバシステムにおいて、前記ディスク装置におけるデータを論理的なファイルとして扱うためのファイル管理データを、予め2つ以上の情報に分割しておき、前記分割された情報を、各サーバ装置内部において管理するファイルシステムを備え、前記複数サーバ装置のうちの1つがこれに接続するクライアントからのデータ配信要求

を受けると、該サーバ装置のファイルシステムは、管理するファイル管理データの情報の中から該当するコンテンツファイルを検索し、該当コンテンツファイルが存在した場合、前記ディスク装置に対してアクセスを行ない、読み出しを行う、ことを特徴とする。

【0009】請求項2に記載のディスク共有型分散サーバシステムは、請求項1に記載のディスク共有型分散サーバシステムにおいて、前記該当コンテンツファイルが存在しなかった場合、前記ファイルシステムは、前記データネットワークを介して接続されている他のサーバ装置に対して、前記該当コンテンツファイルが存在するかどうかを確認し、前記該当コンテンツファイルの情報を持つ前記他のサーバ装置よりデータの読み出しに必要な情報を受け取る手段を備え、前記ディスク装置に対してアクセスを行い、読み出しを行う、ことを特徴とする。

【0010】請求項3に記載のディスク共有型分散サーバシステムは、請求項1または請求項2に記載のディスク共有型分散サーバシステムにおいて、前記複数サーバ装置のうちの1つがこれに接続するクライアントからデータ記録要求を受けると、該サーバ装置のファイルシステムは、前記データネットワークを介して接続されている、他のサーバ装置が管理するファイル管理データの情報の中に、記録要求されているコンテンツファイルと同一名のものがない事を確認する手段と、自分が管理するファイル管理データの情報の中から、記録するのに必要なディスク容量を確保する手段を備え、前記ディスク装置に対して書き込みを行う、ことを特徴とする。

【0011】請求項4に記載のディスク共有型分散サーバシステムは、請求項3に記載のディスク共有型分散サーバシステムにおいて、前記ファイルシステムがファイル管理データを参照して記録するのに必要なディスク容量が確保できなかった場合、前記ファイルシステムは、前記データネットワークを介して接続されている、他のサーバ装置内部にあるファイルシステムに対してコンテンツファイルの情報の管理を要求する手段と、前記ディスク装置に対して書き込む際に、前記コンテンツファイルの情報の管理を要求した他のサーバ装置よりファイルアクセス（書き込み）に必要な情報を受け取る手段を備え、前記ディスク装置に対して書き込みを行う、ことを特徴とする。

【0012】請求項5に記載のディスク共有型分散サーバシステムは、請求項3または請求項4に記載のディスク共有型分散サーバシステムにおいて、各サーバ装置内部のファイルシステムは、共有している前記ディスク装置に記録されている全コンテンツファイル名および前記コンテンツファイルの情報を管理しているサーバ装置を対応付けさせた管理テーブルを持つ、ことを特徴とする。

【0013】請求項6に記載のディスク共有型分散サーバシステムは、請求項5に記載のディスク共有型分散サーバシステムにおいて、コンテンツファイルの追加または

削除を行ったサーバ装置は、前記データネットワークを介して接続されている他のサーバ装置に対して、前記管理テーブルの更新を指示する手段を備えた、ことを特徴とする。

【0014】請求項7に記載のディスク共有型分散サーバシステムは、請求項5または請求項6に記載のディスク共有型分散サーバシステムにおいて、各サーバ装置内部のファイルシステムは、前記分割した記録領域における空き情報を空き領域テーブルとして持つ、ことを特徴とする。

【0015】請求項8に記載のディスク共有型分散サーバシステムは、請求項7に記載のディスク共有型分散サーバシステムにおいて、前記空き領域情報の更新を行ったサーバ装置は、前記データネットワークを介して接続されている他のサーバ装置に対して、前記空き領域テーブルの更新を指示する手段を備えた、ことを特徴とする。

【0016】請求項9に記載のディスク共有型分散サーバシステムは、請求項3または請求項4に記載のディスク共有型分散サーバシステムにおいて、システム初期化時に、前記ディスク装置における記録領域を論理的に分割する手段と、前記分割した記録領域に内在するコンテンツファイルの情報を取得する手段を備えた、ことを特徴とする。

【0017】上記のような構成によって、各ファイルの管理データ、コンテンツテーブル、空き領域の情報を複数のサーバ装置に分散させたことにより、1つのサーバ装置に障害が発生した場合でも、他のサーバ装置が管理するディスクエリアのコンテンツファイルの入出力に関しては支障なく動作することが可能となる。また、システム構成を変更した場合でも、システムを再立ち上げすることにより、他サーバ装置のディスク情報を容易に持つことができる、という作用を有する。

【0018】

【発明の実施の形態】以下、本発明にかかるディスク共有型分散サーバシステムの実施例について、図を用いて説明する。

（実施の形態1）以下に本発明の請求項1ないし請求項8に対応する実施の形態1について説明する。図1は、本発明の実施の形態1におけるディスク共有型分散サーバシステムの構成図を示したものである。図1では3つのサーバ装置が接続されているシステムを例としている。

【0019】図中、11a、11b、11cは、クライアントからビデオデータなどの記録再生要求を受け取って動作するサーバ装置A、B、Cを示し、ハードウェア構成およびソフトウェア構成を全く同じとする。13はビデオデータを含むコンテンツが格納されている共有ディスク装置である。各サーバ装置11a、11b、11cと共有ディスク装置13とは、Fibre Channelに代表されるような複数のプロトコルによるデー

タの送受信が可能なデータネットワーク14で接続されている。また、各サーバ装置11a、11b、11c内部のファイルシステムは、共有ディスク装置13上に格納したデータの集合を論理的なファイルとして扱うため、それぞれファイル管理データA12a、ファイル管理データB12b、ファイル管理データC12cを持っており、共有ディスク装置13に対してアクセスすることが可能になっている。

【0020】次にファイル管理データの構成を図2において示す。図中、12a、12b、12cは、各サーバ装置内部のファイルシステムが管理しているファイル管理データA、B、Cのイメージ図である。各ファイル管理データ12a、12b、12cは、主に2つの情報から構成されている。1つ目は共有ディスク装置13の記録領域の状態（使用／未使用）を示すディスク領域情報であり、共有ディスク装置13を論理的に3つのエリアA、B、Cに分割し、それぞれ管理エリアAのディスク領域情報21a、管理エリアBのディスク領域情報21b、管理エリアCのディスク領域情報21cで示している。これらのディスク領域情報は互いにエリア重複していない。2つ目はディスク領域情報22a、22b、22cに示されている記録領域内に存在しているコンテンツの情報であり、22a、22b、22cはそれぞれ、エリアAに含まれているコンテンツファイル1～p情報、エリアBに含まれているコンテンツファイルp+1～q情報、エリアCに含まれているコンテンツファイルq+1～r情報を示している。

【0021】次に図1の構成において、システムにおける動作を、図3及び図4を参照して説明する。各図において、太字破線矢印は制御またはデータアクセス、太字実線矢印はデータの流れを示す。図3は、サーバ装置がこれに接続されているクライアントからコンテンツファイルm（ $1 \leq m \leq p$ ）の配信要求を受けた場合の動作を示す説明図である。サーバ装置A11aが接続されているクライアントからコンテンツファイルm（ $1 \leq m \leq p$ ）の配信要求31を受け取ったとする。サーバ装置A11aのファイルシステムは、ファイル管理データA12aを参照してコンテンツファイルmの情報検索32を行う。該当したコンテンツファイルm情報を元に共有ディスク装置13からデータ読み出し33を行ない、所定の出力先へデータ配信34を行う。

【0022】図4は、サーバ装置がこれに接続されているクライアントからコンテンツファイルn（ $p+1 \leq n \leq q$ ）の配信要求を受けた場合の動作を示す説明図である。サーバ装置A11aが接続されているクライアントからコンテンツファイルn（ $p+1 \leq n \leq q$ ）の配信要求41を受け取ったとする。サーバ装置A11aのファイルシステムはファイル管理データA12aを参照してコンテンツファイルnの情報検索42を行うが、該当情報が見つからないので、サーバ装置B11bに対してデ

ータネットワーク14を通じて情報検索の依頼43を行う。サーバ装置B11bのファイルシステムはファイル管理データB12bを参照してコンテンツファイルnの情報検索44を行う。該当したコンテンツファイルn情報をサーバ装置A11aに送信45することによりサーバ装置A11aは共有ディスク装置13からデータの読み出し46を行い、所定の出力先へデータ配信47を行う。上記の方法の場合、コンテンツファイルnの検索を、接続されているサーバ装置全部に対して問い合わせなければならない状況が発生するため、さらに効率よく検索できるよう、図5に示すようなテーブルを用意する。

【0023】図5は、コンテンツテーブルの内容を示す説明図である。図5に示すように、コンテンツテーブルはコンテンツファイル名に相対して、状態および保存場所を示している。これは図1で示した構成のサーバ装置A、B、Cにおいて、各々が同期を取って同じ内容のテーブルを持つこととする。このテーブルの更新（コンテンツファイルの追加および削除）を行うサーバ装置は、他のサーバ装置に対して情報の更新を指示することで、データ同期を保つ。

【0024】図6は、図1の構成に図5で示したコンテンツテーブルを備えたディスク共有型分散サーバシステムの構成を示す説明図である。図6において、11a、11b、11cはサーバ装置、12a、12b、12cはファイル管理データ、13は共有ディスク装置、14はデータネットワークである。また、60a、60b、60cは、それぞれサーバ装置11a、11b、11cに含まれるコンテンツテーブルである。次に図6の構成において、システムにおける動作を、図7、図8及び図9を参照して説明する。各図において、太字破線矢印は制御またはデータアクセス、太字実線矢印はデータの流れを示す。

【0025】図6は、サーバ装置がこれに接続されているクライアントからコンテンツファイルn ($p+1 \leq n \leq q$) の配信要求を受けた場合の動作を示す説明図である。サーバ装置A11aが接続されているクライアントからコンテンツファイルn ($p+1 \leq n \leq q$) の配信要求61を受け取ったとする。サーバ装置A11aのファイルシステムはコンテンツテーブル60aを元にして該当コンテンツファイルがサーバ装置B11bの管理下であることを確認し、サーバ装置B11bに対してデータネットワーク14を通じて情報検索の依頼62を行う。サーバ装置B11bのファイルシステムは、ファイル管理データB12bを参照してコンテンツファイルnの情報取得63を行い、ファイルn情報をサーバ装置A11aに送信64することにより、サーバ装置A11aは共有ディスク装置13からデータの読み出し65を行い、所定の出力先へデータ配信66を行う。

【0026】図7は、図5で示したコンテンツテーブル

を備えたサーバ装置が、これに接続されているクライアントから、新しいコンテンツファイルxの記録要求を受けた場合の動作を示す説明図である。図7において、サーバ装置A11aが接続されているクライアントから新しいコンテンツファイルxの記録要求71を受け取ったとする。サーバ装置A11aのファイルシステムはコンテンツテーブル60aにおいて同一のコンテンツファイル名が存在しない事を確認した後、ファイル管理データA12aにおいて、共有ディスク装置13の記録領域エリアAに空き領域が存在する事を確認して、新しいコンテンツファイルxをコンテンツテーブル60aに追加する。次に所定の入力先よりデータ受信73を行ない、記録場所をファイル管理データA12aにおけるコンテンツファイルxの情報に書き込み74、共有ディスク装置13にデータ書き込み75を行う。

【0027】図8は、図5で示したコンテンツテーブルを備えたサーバ装置が、これに接続されているクライアントから新しいコンテンツファイルの記録要求を受け、他のサーバ装置のファイル管理データを用いる場合の動作を示す説明図である。図8において、サーバ装置A11aが接続されているクライアントから新しいコンテンツファイルxの記録要求81を受け取ったとする。サーバ装置A11aのファイルシステムはコンテンツテーブルにおいて同一のコンテンツファイル名が存在しない事を確認した後、ファイル管理データA12aにおいて、共有ディスク装置13の記録領域エリアAに空き領域が存在しないと判断82した場合、サーバ装置B11bに対してデータネットワーク14を通じてエリアBの空き領域の確認依頼83を行う。サーバ装置B11bのファイルシステムは、ファイル管理データB12bを参照して、エリアBの空き領域の確認84を行うと共に、新しいコンテンツファイルxをコンテンツテーブル60bに追加する。次に空き領域情報つまり共有ディスクにおいて書き込み可能な物理領域の情報をサーバ装置A11aに送信85することにより、サーバ装置A11aは所定の入力先よりデータ受信86を行い、記録場所をファイル管理データB12bにおけるコンテンツファイルxの情報に書き込み87、共有ディスク装置13にデータ書き込み88を行う。

【0028】図9は、図1の構成に図5で示したコンテンツテーブルと、後述する領域テーブルとを備えたディスク共有型分散サーバシステムの構成を示す説明図である。図9において、11a、11b、11cはサーバ装置、12a、12b、12cはファイル管理データ、13は共有ディスク装置、14はデータネットワーク、60a、60b、60cはコンテンツテーブルである。また、90a、90b、90cは、共有ディスク装置13における空き容量をエリア別に区分している領域テーブルである。これは図1で示した構成のサーバ装置A、B、Cにおいて、各々が同期を取って同じ内容のテーブ

ルを持つこととする。このテーブルの更新（空き領域の増減）を行うサーバ装置は、他のサーバ装置に対して情報の更新を指示することで、データ同期を保つ。

【0029】次に動作について図9を参照しながら説明する。図において、太字破線矢印は制御またはデータアクセス、太字実線矢印はデータの流れを示す。サーバ装置A11aに接続されているクライアントから新しいコンテンツファイルxの記録要求91を受け取ったとする。サーバ装置A11aのファイルシステムは、図5で示したコンテンツテーブルにおいて、同一のコンテンツファイル名が存在しない事を確認した上で、領域テーブル90aを参照し、予め設定された優先度に従って空き領域のあるエリアを確定し、そのエリアを管理するサーバ装置を特定する。仮にそれがサーバ装置B11bであった場合、データネットワーク14を通じて空き領域情報の要求92を送信し、サーバ装置B11bは空き領域情報を取得93した上でサーバ装置A11aに情報を送信94する。サーバ装置A11aは所定の入力先よりデータ受信95を行い、記録場所をファイル管理データB12bにおけるコンテンツファイルxの情報に書き込む96と同時に、共有ディスク装置13にデータ書き込み97を行う。

【0030】以上のように、本発明の実施の形態1に係るディスク共有型分散サーバシステムによれば、システム固有の排他制御を司る機関の存在しない本ディスク共有型分散サーバシステムにおいて、各ファイルの管理データを複数のサーバ装置に分散させ、複数のサーバ装置が同期を保っているコンテンツテーブルと、領域テーブルとを各サーバ装置に備えたことにより、1つのサーバ装置に障害が発生した場合、該サーバ装置が管理するディスクエリアに記録されているコンテンツファイルの入出力は停止してしまうが、他のサーバ装置が管理するディスクエリアのコンテンツファイルの入出力に関しては全く支障無く動作することが可能となり、コンテンツファイルまたは空き領域の情報の検索を、接続されているサーバ装置全部に対して問い合わせすることなく行うことができる。

【0031】（実施の形態2）以下に本発明の請求項9に対応する実施の形態2について説明する。図10は、本発明の実施の形態2に係るディスク共有型分散サーバシステムの構成を示したものである。図中、101-1～101-nは、ディスク一体型のサーバ装置である。これらサーバ装置およびディスク装置101-1～101-nは、Fibre Channelに代表されるような複数のプロトコルによるデータの送受信が可能なデータネットワーク102で接続されている。また、各サーバ装置101-1～101-n内部のファイルシステムは、一体となっているディスク装置上に格納したデータの集合を論理的なファイルとして扱うためのファイル管理データを、それぞれが持っている。

【0032】この構成における動作について以下に説明する。システム起動時、つまりサーバ装置およびディスク装置101-1～101-nの全ての電源が投入された状態において、各サーバ装置は、データネットワーク102に接続されている他のサーバ装置に対して、各々が管理しているコンテンツの一覧、及び各々が管理しているディスク容量、の2つについて問い合わせを行う。問い合わせを受けたサーバ装置は、ファイル管理テーブルよりその情報を取得して返答する。全ての返答を受け取ると、図5で示したようなコンテンツテーブル、および図9で示した領域テーブルを構築する事で、他のサーバ装置のディスク情報を持つ事になる。

【0033】以上のように、本発明の実施の形態2に係るディスク共有型分散サーバシステムによれば、上記のような初期化手順を持つ事により、障害が発生した後にシステム構成を変更した場合でも、システムを再立ち上げすることにより、他サーバ装置のディスク情報を容易に持つことが可能である。また、ディスク容量の拡張や、入出力ポートの拡張などの目的により、サーバ装置およびディスク装置の台数を増やす場合でも、システムを再立ち上げすることにより、他サーバ装置のディスク情報を容易に持つことが可能である。

【0034】

【発明の効果】以上のように、本発明の請求項1および請求項2に係るディスク共有型分散サーバシステムによれば、データの入出力を目的とした複数サーバ装置が、データネットワークを介して、データを格納するためのディスク装置を共有する構成をとるディスク共有型分散サーバシステムにおいて、前記ディスク装置におけるデータを論理的なファイルとして扱うためのファイル管理データを、予め2つ以上の情報に分割しておき、前記分割された情報を、各サーバ装置内部において管理するファイルシステムを備え、前記複数サーバ装置のうちの1つがこれに接続するクライアントからのデータ配信要求を受けると、該サーバ装置のファイルシステムは、管理するファイル管理データの情報の中から該当するコンテンツファイルを検索し、該当コンテンツファイルが存在した場合、前記ディスク装置に対してアクセスを行ない、また、前記該当コンテンツファイルが存在しなかった場合、前記ファイルシステムは、前記データネットワークを介して接続されている他のサーバ装置に対して、前記該当コンテンツファイルが存在するかどうかを確認し、前記該当コンテンツファイルの情報を保持する前記他のサーバ装置よりデータの読み出しに必要な情報を受け取る手段を備え、前記ディスク装置に対してアクセスを行い、読み出しを行うので、1つのサーバ装置に障害が発生した場合でも、該サーバ装置が管理するディスクエリアに記録されているコンテンツファイルの読み出し処理は停止するものの、他のサーバ装置が管理するディスクエリアのコンテンツファイルの読み出しに関しては全く

支障無く動作することが可能となる、という効果を得られる。

【0035】また、請求項3および請求項4に係るディスク共有型分散サーバシステムによれば、請求項1または請求項2記載のディスク共有型分散サーバシステムにおいて、前記複数サーバ装置のうちの1つがこれに接続するクライアントからデータ記録要求を受けると、該サーバ装置のファイルシステムは、前記データネットワークを介して接続されている、他のサーバ装置が管理するファイル管理データの情報の中に、記録要求されているコンテンツファイルと同一名のものがない事を確認する手段と、前記ファイルシステムが管理するファイル管理データの情報の中から、記録するのに必要なディスク容量を確保する手段を備え、また、前記ファイルシステムがファイル管理データを参照して記録するのに必要なディスク容量が確保できなかった場合、前記ファイルシステムは、前記データネットワークを介して接続されている、他のサーバ装置内部にあるファイルシステムに対してコンテンツファイルの情報の管理を要求する手段と、前記ディスク装置に対して書き込む際に、前記コンテンツファイルの情報の管理を要求した他のサーバ装置よりデータの書き込みに必要な情報を受け取る手段を備え、前記ディスク装置に対して書き込みを行うので、1つのサーバ装置に障害が発生した場合でも、該サーバ装置が管理するディスクエリアへのコンテンツファイルの記録処理は停止するものの、他のサーバ装置が管理するディスクエリアへの記録処理に関しては全く支障無く動作することが可能となる、という効果を得られる。

【0036】また、請求項5および請求項6に係るディスク共有型分散サーバシステムによれば、請求項3または請求項4記載のディスク共有型分散サーバシステムにおいて、各サーバ装置内部のファイルシステムは、共有している前記ディスク装置に記録されている全コンテンツファイル名および前記コンテンツファイルの情報を管理しているサーバ装置を対応付けさせた管理テーブルを持ち、また、コンテンツファイルの追加または削除を行ったサーバ装置は、前記データネットワークを介して接続されている他のサーバ装置に対して、前記管理テーブルの更新を指示する手段を備えたので、コンテンツファイルの整合性を保ち、コンテンツファイル情報の検索を、接続されているサーバ装置全部に対して問い合わせることなく、効率良くデータを検索することができる、という効果を得られる。

【0037】また、請求項7および請求項8に係るディスク共有型分散サーバシステムによれば、請求項5または請求項6記載のディスク共有型分散サーバシステムにおいて、各サーバ装置内部のファイルシステムは、前記分割した記録領域における空き情報を空き領域テーブルとして持ち、また、前記空き領域情報の更新を行ったサーバ装置は、前記データネットワークを介して接続され

ている他のサーバ装置に対して、前記空き領域テーブルの更新を指示する手段を備えたので、ファイルシステム間における空き領域の整合性を保ち、接続されているサーバ装置全部に対して問い合わせをすることなく、効率良く空き領域の情報を検索することができる、という効果を得られる。

【0038】また、請求項9に係るディスク共有型分散サーバシステムによれば、請求項3または請求項4記載のディスク共有型分散サーバシステムにおいて、システム初期化時に、前記ディスク装置における記録領域を論理的に分割する手段と、前記分割した記録領域に内在するコンテンツファイルの情報を取得する手段を備えたので、障害復旧時やサーバ装置およびディスク装置を増やすなどの場合に、各サーバ装置内部においてのファイルシステム情報の更新を容易にする、という効果を得られる。

【図面の簡単な説明】

【図1】本発明の実施の形態1による、ディスク共有型分散サーバシステムを示す構成図。

【図2】本発明の実施の形態1による、図1に示すディスク共有型分散サーバシステムに備えられるサーバ装置内部のファイルシステムが扱うファイル管理データの概要図。

【図3】本発明の実施の形態1による、ディスク共有型分散サーバシステムのデータ配信動作一例を示した図。

【図4】本発明の実施の形態1による、ディスク共有型分散サーバシステムのデータ配信動作一例を示した図。

【図5】本発明の実施の形態1による、図1に示すディスク共有型分散サーバシステムに備えられるサーバ装置内部のファイルシステムが扱うコンテンツテーブルの概要図。

【図6】本発明の実施の形態1による、ディスク共有型分散サーバシステムのデータ配信動作の一例を示した図。

【図7】本発明の実施の形態1による、ディスク共有型分散サーバシステムのデータ記録動作の一例を示した図。

【図8】本発明の実施の形態1による、ディスク共有型分散サーバシステムのデータ記録動作の一例を示した図。

【図9】本発明の実施の形態1による、ディスク共有型分散サーバシステムのデータ記録動作の一例を示した図。

【図10】本発明の実施の形態2による、ディスク共有型分散サーバシステムを示す構成図。

【符号の説明】

11a, 11b, 11c…サーバ装置

12a, 12b, 12c…ファイル管理データ

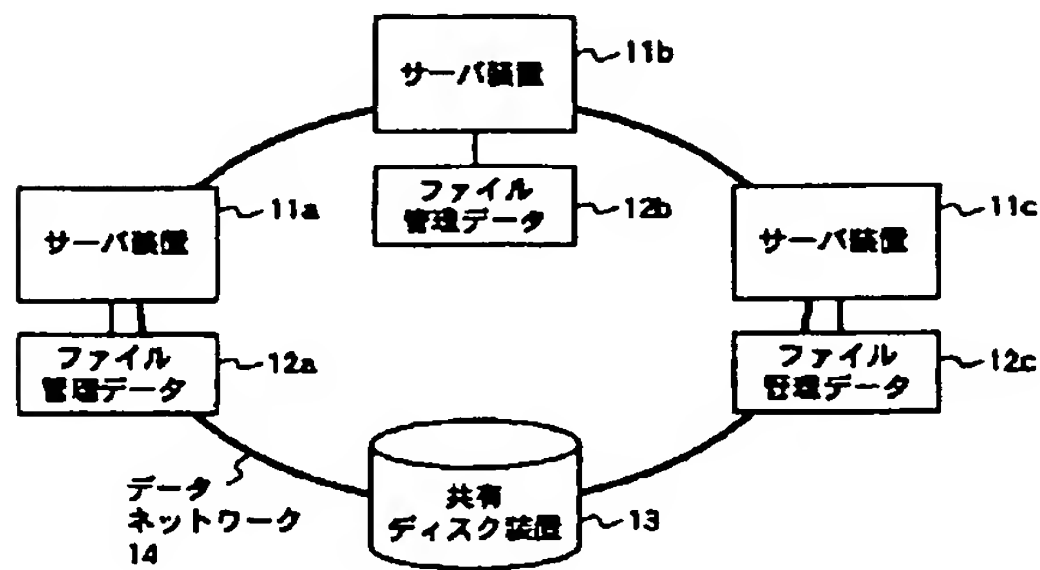
13…共有ディスク装置

14, 102…データネットワーク

13

21a, 21b, 21c…ディスク領域情報
 22a, 22b, 22c…コンテンツファイル情報
 31, 32, 41, 42, 43, 44, 45, 61, 62, 63, 64, 71, 72, 74, 81, 82, 83, 84, 85, 87, 91, 92, 93, 94, 96
 …制御信号

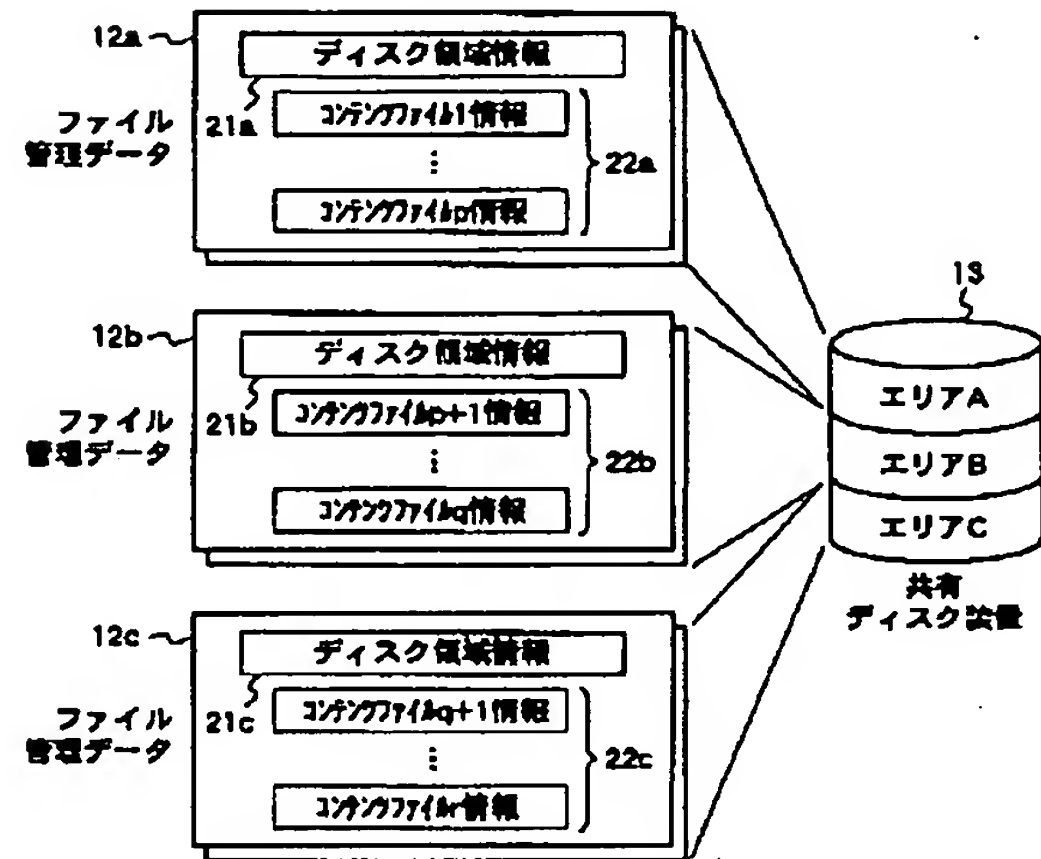
【図1】



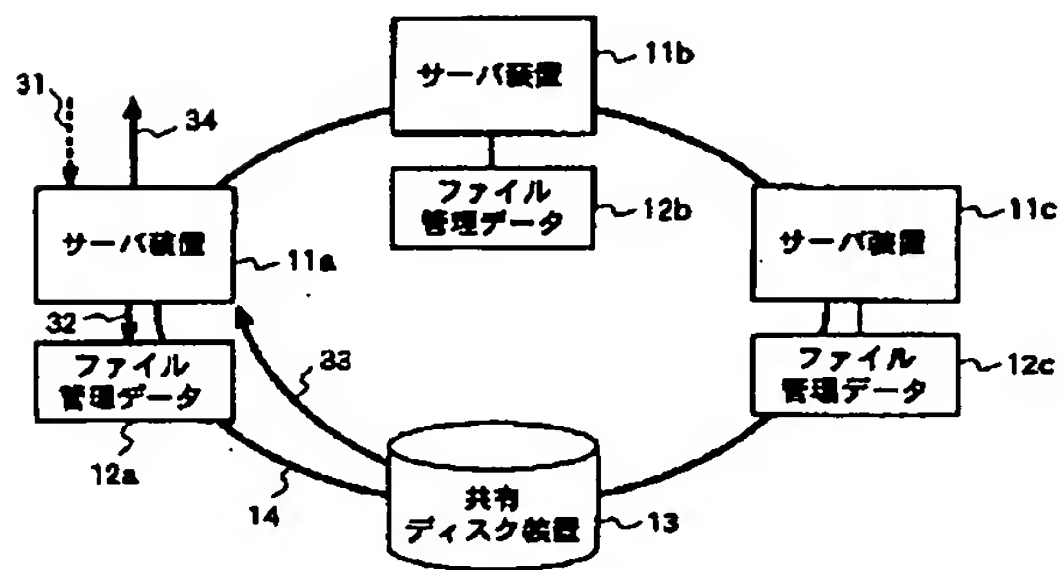
14

33, 34, 46, 47, 65, 66, 73, 75, 86, 88, 95, 97…データの流れ
 60a, 60b, 60c…コンテンツテーブル
 90a, 90b, 90c…領域テーブル
 101-1~101-n…ディスク型サーバ装置

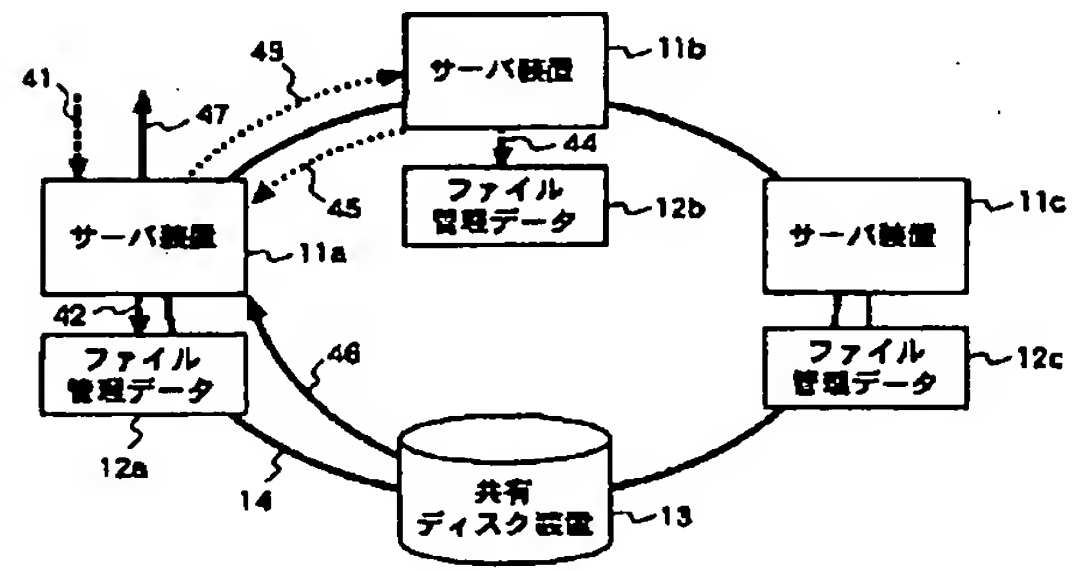
【図2】



【図3】



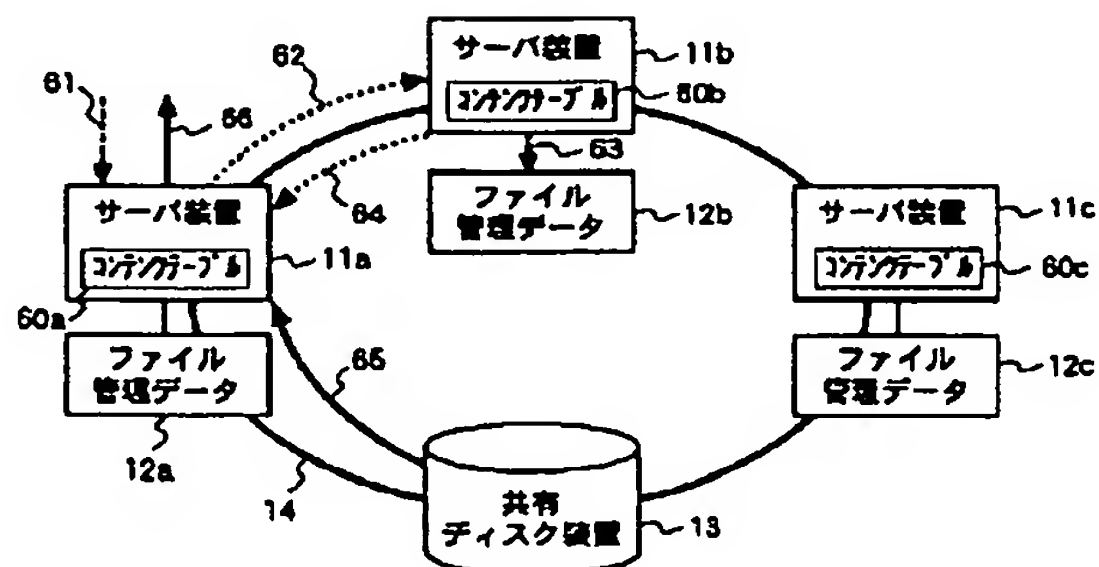
【図4】



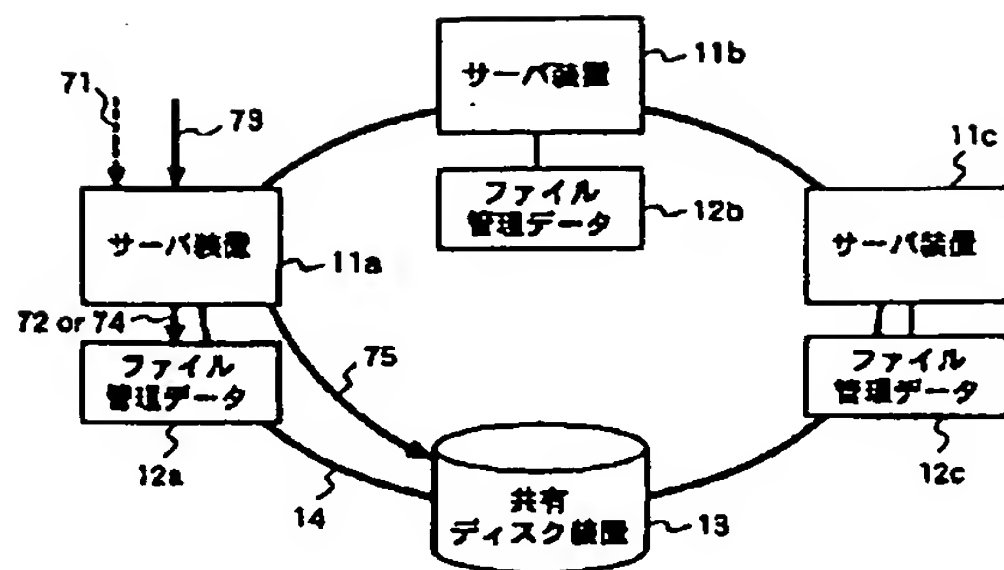
【図5】

コンテンツファイル名	状態	保存場所
コンテンツファイル1	未使用	エリアA
...		
コンテンツファイルp	記録中	
コンテンツファイルp+1		エリアB
...		
コンテンツファイルq		
コンテンツファイルq+1		エリアC
...		
コンテンツファイルr		

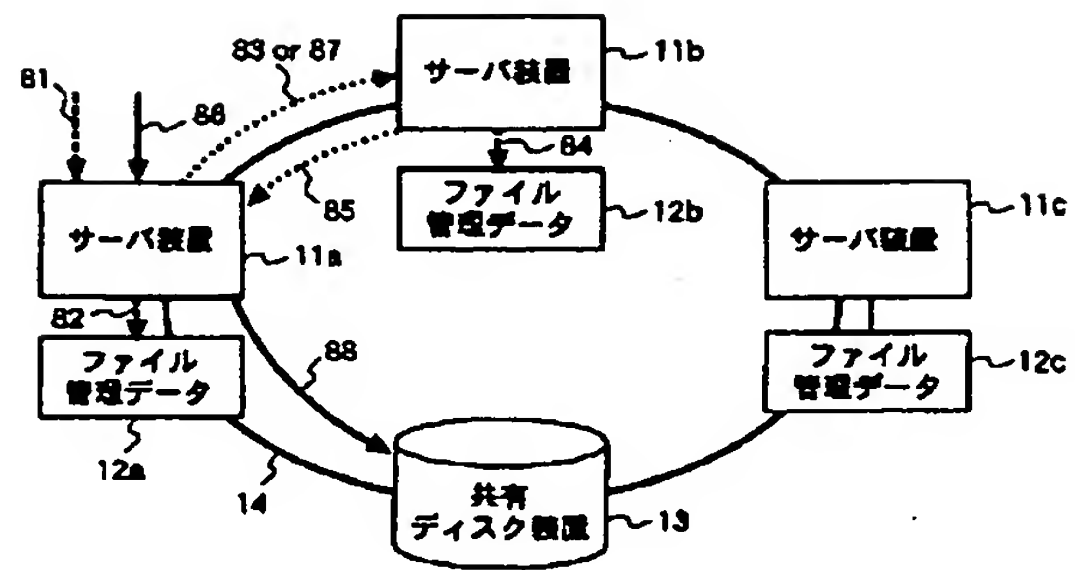
【図6】



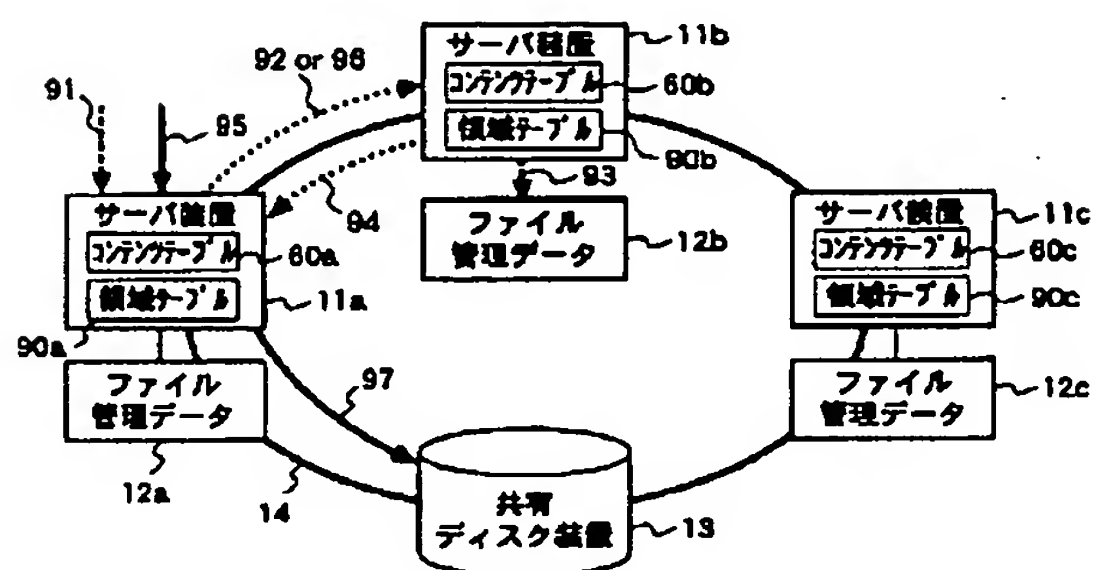
【図7】



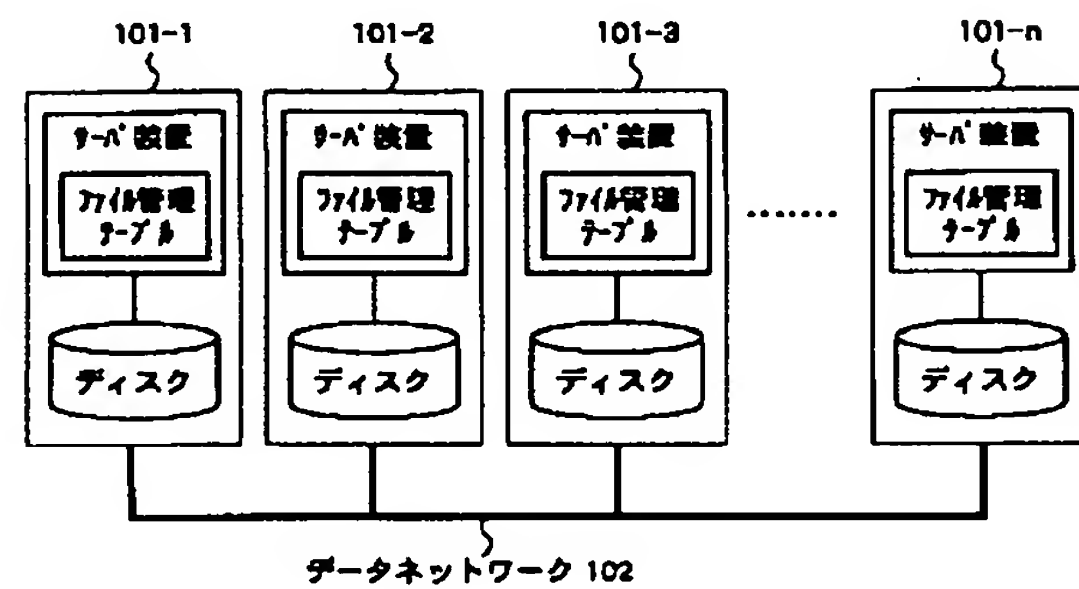
【図8】



【図9】



【図10】



フロントページの続き

(51) Int. Cl.⁷
G 0 6 F 15/167

識別記号

F I
G 0 6 F 15/167

テーマコード (参考)
B